

Metadata Thoughts & Lessons Learned From Digital Libraries

Katy Ginger (ginger@ucar.edu)

Meteorologist & Metadata Architect

Digital Learning Sciences (DLS)

University Corporation for Atmospheric Research (UCAR)

A Better Metadata Definition

- Not just data about data
- Metadata can also be about applications, networks or services
- *Data that describes any aspect of an enterprise's information assets and enables the use and management of these assets in meaningful ways.*

Trends in Asset Management

- Integrate existing assets
- Create and collect new assets only when needed
- Integration requires detailed understanding of asset sources, uses and owners
- Asset integration brings the expectation of services to manage, extract, transform, verify, distribute/make accessible assets (interoperability ideas)
- *Metadata is critical to all these functions.*

Trends in Earth Data Management

- Academic research community & NSF:
 - Two awards to: John Hopkins Univ (with **UCAR as co-PI**) and Univ. of New Mexico
 - Create a **Data Conservancy** for data integration & curation
- International data managers:
 - Lack of data integration and curation is hampering science (OCED - <http://scholarship20.blogspot.com/2009/04/paris-20th-april-2009-oecd-has-released.html>)
 - Need the ability to discover appropriate and useful data such that a group (includes UCAR) is investigating the use of DOI (digital object identifiers) for datasets

Broad Metadata Types

- **Administrative**: information on the management of an asset -- like ownership, terms of use, intellectual property & preservation
 - (e.g. private mesonet data ownership and terms of use)
- **Descriptive**: describes an asset for purposes of discovery, use and identification
 - (e.g. hourly observations; instrument description)
- **Structural**: indicates how assets are put together to create a greater whole or service
 - (e.g. stations that comprise a mesonet network)

Metadata Components

- **Structure**: the metadata fields & the enforcement of valid information (e.g. XML schemas & metadata records)
- **Values**: the content of metadata fields generated manually, automatically or using certain terms (controlled vocabularies) (e.g. type of data: sfc or upper air; or a temperature value)
- **Cataloging Rules**: best practices on how to generate appropriate metadata (e.g. date must be present and in the format yyyy-mm-dd)
- **Sharing**: methods for the provision of metadata to applications, services & end users (e.g. web services)

Two Metadata Stories

1. Digital Library for Earth System Education (DLESE)

- Earth science focused
- Educationally oriented (K-12 slant)
- Asked to address datasets

2. National Science Digital Library (NSDL)

- Science, Engineering, Technology, Math and Medical (STEM) focused (higher ed slant)
- Topical contributors

Two Metadata Stories

DLESE

- **Metadata framework:** Based on IMS because of education focus & geospatial
- **Metadata framework:** 4 metadata frameworks work with 4 different types of assets ([learning objects](#), [collection information](#), [annotations](#), [news/events](#))
- **Required metadata:** 10 fields with strong data typing enforced and controlled vocabularies ([e.g. title](#), [URL](#), [description](#), [resource type](#))

NSDL

- **Metadata framework:** Qualified Dublin Core & four IEEE-LOM fields
- **Metadata framework:** 1 metadata framework for all assets types. Cannot tell the difference nor find learning objects versus annotations
- **Required metadata:** Only URL is required; has some controlled vocabularies

Two Metadata Stories (cont'd)

DLESE

- **Browsing:** Comprehensive as a result of controlled vocabularies and metadata frameworks
- **Crosswalking to:** Harder to map existing metadata to
- **Crosswalking from:** Easy to map to other metadata because DLESE has strong data typing and a set of required metadata
- **Geospatial:** Defines geospatial & temporal info for use with Google Earth

NSDL

- **Browsing:** Weak browse structure at collection-level
- **Crosswalking to:** Easier to map existing metadata to
- **Crosswalking from:** Sometimes easy, sometimes hard to map to other metadata depending on what you are mapping to because few NSDL required metadata
- **Geospatial:** Does not define how geospatial information is entered; Not good for Google Earth

Two Metadata Stories (cont'd)

DLESE

- **Sharing:** Groups can use a variety of methods to send metadata
- **Metadata creation support:** offers a web-based cataloging tool that includes a metadata sharing mechanism and the ability to write to the NSDL metadata storage API
- **Metadata framework change support:** In the past, provided extensive support

NSDL

- **Sharing:** Requires groups to use OAI to send metadata but new methods will soon exist
- **Metadata creation support:** Now offers the DLESE web-based cataloging tool
- **Metadata framework change support:** Provides some support

Lessons Learned: Overview

- Metadata Frameworks
- Vocabularies
- Metadata Creation
- Storage Structures
- Catalog Tools
- Contributing

Lessons Learned: Frameworks 1

- No required metadata results in very little metadata
- Use metadata appropriate to the asset being managed (improves metadata quality & supports discovery)
- Appropriate metadata may or may not involve the use of existing metadata framework standards (e.g. [FGDC](#), [Dublin Core](#), [IEEE-LOM](#), [ISO](#))
- Use metadata fields and enforcement mechanisms that are reasonably flexible (change is inevitable)

Lessons Learned: Frameworks 2

- Be explicit on accepted metadata versions & formats
- Are multiple metadata formats allowed or only one format for a particular type of asset?
- Use a metadata framework that can be mapped to other metadata frameworks without much loss of data
- More organizations/people (than you can think of) will ask to share metadata with you or request your metadata for their project
- Know the copyright & terms of use of the metadata shared and ingested

Lessons Learned: Vocabularies

- Extremely useful in creating browse capabilities
- Useful in knowing what's in a repository (or network)
- Definitions of vocabulary terms are necessary to provide meaning
- Promotes greater metadata consistency
- Decide how terms are managed and aged off
- Decide if a vocabulary registry will be used (terms become URLs then)
- Decide if vocabularies will be required and enforced

Lessons Learned: Metadata Creation

- Does nature of community require
 - Tools to support metadata creation
 - Metadata and metadata aggregation training
- Support legacy/existing metadata formats semantically and technically:
 - Help map vocabularies
 - Help change metadata formats programmatically
- Require metadata that is of extremely high value

Lessons Learned: Storage Structures

- Metadata is just one piece of information for an asset
- Choose flexible storage structures for metadata, content indexing, content storage and user supplied notes/annotations (e.g. digital libraries use Fedora repositories and Lucene indexes)
- Some metadata information is best stored locally (e.g. sensor site) and some is best stored centrally (e.g. metadata about an instrument); then decide how and when to access centralized metadata

Lessons Learned: Catalog Tools

- Make flexible – DLESE Collection System (DCS) generates user interfaces and vocabs directly from XML schemas that define a metadata framework
- Provide metadata generation support by
 - Defining the terms of each controlled vocabulary used
 - Providing cataloging best practices for cataloging each field
- Have built-in metadata sharing capabilities (OAI, API, web services)

Lessons Learned: Contributing

- Allow contributors to provide metadata via multiple methods (OAI, APIs, web services and many more)
- Contributors are more inclined to contribute content (data) rather than metadata
- If willing to contribute metadata, often not aware of metadata format to be contributed or metadata best practices
- Don't understand the relationship between metadata and its ability to support uniform or targeted discovery

Focus on Metadata Creation Part

- Getty Museum: ***Practical Principles for Metadata Creation and Maintenance***
http://www.getty.edu/research/conducting_research/standards/intrometadata/principles.pdf

More Resources

- DLESE metadata: <http://www.dlese.org/Metadata>
- Cataloging tool: <http://ncs.nsd.org/mgr> Try as a **GUEST**
- Controlled vocabulary terminology & concepts:
http://www.dlese.org/Metadata/vocabularies/term_expln.php
- DLESE & Google Earth: <http://search.dlese.org/maps/>
- OAI-PMH software: jOAI (sharing metadata)
http://www.dlese.org/dds/services/joai_software.jsp